Granger causality in dynamic binary short panel data models

Francesco Bartolucci and Claudia Pigini

University of Perugia, Marche Polytechnic University

13 March 2017

Granger causality in dynamic binary short panel data models

Francesco Bartolucci
Università di Perugia (IT)
francesco.bartolucci@unipg.it

Claudia Pigini
Università Politecnica delle Marche and MoFiR (IT)
c.pigini@univpm.it

March 13, 2017

Abstract

Strict exogeneity of covariates other than the lagged dependent variable, and conditional on unobserved heterogeneity, is often required for consistent estimation of binary panel data models. This assumption is likely to be violated in practice because of feedback effects from the past of the outcome variable on the present value of covariates and no general solution is yet available. In this paper, we provide the conditions for a logit model formulation that takes into account feedback effects without specifying a joint parametric model for the outcome and predetermined explanatory variables. Our formulation is based on the equivalence between Granger's definition of noncausality and a modification of the Sims' strict exogeneity assumption for nonlinear panel data models, introduced by Chamberlain (1982) and for which we provide a more general theorem. We further propose estimating the model parameters with a recent fixed-effects approach based on pseudo conditional inference, adapted to the present case, thereby taking care of the correlation between individual permanent unobserved heterogeneity and the model's covariates as well. Our results hold for short panels with a large number of cross-section units, a case of great interest in microeconomic applications.

Keywords: fixed effects, noncausality, predetermined covariates, pseudo-conditional inference, strict exogeneity.

JEL Classification: C12, C23, C25

1 Introduction

There is an increasing number of empirical microeconomic applications that require the estimation of binary panel data models, which are typically dynamic so as to account for state dependence (Heckman, 1981).[1] In these contexts, strict exogeneity of covariates other than the lagged dependent variable, conditional on unobserved heterogeneity, is required for consistent estimation of the regression and state dependence parameters, when the estimation relies on correlated random effects or on fixed effects which are eliminated when conditioning on suitable sufficient statistics for the individual unobserved heterogeneity. However, the assumption of strict exogeneity is likely to be violated in practice because there may be feedback effects from the past of the outcome variable on the present values of the covariates, namely the model covariates may be Granger-caused by the response variable Granger (1969). While in linear models the mainstream approach to overcome this problem is to consider instrumental variables (Anderson and Hsiao, 1981; Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998), considerably fewer results are available for nonlinear binary panel data models with predetermined covariates. This is particularly true with short binary panel data and no general solution is yet available, despite the relevance of binary these type of data in microeconomic applications.

Honoré and Lewbel (2002) propose a semiparametric estimator for the parameters of a binary choice model with predetermined covariates. However, they provide identification conditions when there is a further regressor that is continuous, strictly exogenous, and independent of the individual specific effects. These requirements are often difficult to be fulfilled in practice. Arellano and Carrasco (2003) develop a semiparametric strategy based on the Generalized Method of Moments (GMM) estimator involving the probability distribution of the predetermined covariates (sample cell frequencies for discrete covariates or nonparametric smoothed estimates for continuous covariates) that can, however, be difficult to employ when the set of relevant explanatory variables is large. A different approach is taken by Wooldridge (2000), who proposes to specify a joint model for the response variable and the predetermined covariates; the model parameters are estimated by a correlated random-effects approach (Mundlak, 1978; Chamberlain, 1984), to account for the dependence between strictly exogenous explanatory variables and individual unobserved effects, combined with a preliminary version of the Wooldridge (2005)'s

[1] Estimators of dynamic discrete choice models are employed in studies related to labor market participation (Heckman and Borjas, 1980; Arulampalam, 2002; Stewart, 2007), and specifically to female labor supply and fertility choices (Hyslop, 1999; Carrasco, 2001; Keane and Sauer, 2009; Michaud and Tatsiramos, 2011), self-reported health status (Contoyannis et al., 2004; Halliday, 2008; Heiss, 2011; Carro and Traferri, 2012), poverty traps (Cappellari and Jenkins, 2004; Biewen, 2009), welfare participation (Wunder and Riphahn, 2014), unionization of workers (Wooldridge, 2005), household finance (Alessie et al., 2004; Giarda, 2013; Brown et al., 2014), firms' access to credit (Pigini et al., 2016), and migrants' remitting behavior (Bettin and Lucchetti, 2016)

solution to the initial conditions problem. Although this is an intuitive strategy, it relies on distributional assumptions on the individual unobserved heterogeneity; moreover, it is computationally demanding when the number of predetermined covariates is large and it requires strict exogeneity of the covariates used for the parametric random-effects correction.

A strategy similar to that developed by Wooldridge (2000) is adopted by Mosconi and Seri (2006), who test for the presence of feedback effects in binary bivariate time-series by means of Maximum Likelihood (ML)-based test statistics. They build their estimation and testing proposals on the definition of Granger causality (Granger, 1969), which is typical of the time series literature, as adapted to the nonlinear panel data setting by Chamberlain (1982) and Florens and Mouchart (1982). While attractive, Mosconi and Seri's approach does not account for individual time-invariant unobserved heterogeneity and is better suited for quite long panels, whereas applications, such as intertemporal choices related to the labor market, poverty traps, and persistence in unemployment, often rely on very short time-series and a large number of cross-section units resulting from rotated surveys. Furthermore, in the short panel data setting, dealing properly with time-invariant unobserved heterogeneity is crucial for the attainability of the estimation results, since individual-specific effects are often correlated with the covariates of interest. Moreover, the focus is often on properly detecting the causal effects of past events of the phenomenon of interest, namely the *true* state dependence, as opposed to the persistence generated by permanent individual unobserved heterogeneity (Heckman, 1981).

In this paper, we propose a logit model formulation for dynamic binary fixed T-panel data model that takes into account general forms of feedback effects from the past of the outcome variable on the present value of the covariates. Our formulation presents three main advantages with respect to the available solutions. First, it does not require the specification of a joint parametric model for the outcome and predetermined explanatory variables. In fact, the starting point to build the proposed formulation is the definition of noncausality (Granger, 1969), the violation of which corresponds to the presence of feedback effects, as stated in terms of conditional independence by Chamberlain (1982) for nonlinear models. Translating the definition of noncausality to a parametric model requires, however, the specification of the conditional probability for the covariates (x). On the contrary, we follow Chamberlain (1982) and introduce an equivalent definition based on a modification of Sims (1972)'s strict exogeneity for nonlinear models, which only involves specifying the probability for the binary dependent variable at each time occasion (y_t) conditional on past, present, and future values of x, and for which we provide a more general theorem of equivalence to noncausality.

Second, the proposed model has a simple formulation and allows for the inclusion of even a large number of predetermined covariates. Under the logit model, it amounts to augment the linear index function with a linear combination of the leads of the predetermined covariates, along with the lags of the binary dependent variable. We analytically prove that this augmented linear index function corresponds to the logit for the joint distribution of y_t and the future values of x, under the assumption that the distribution of the predetermined covariates belongs to the exponential family with dispersion parameters (Barndorff-Nielsen, 1978) and that their conditional means depend on time-fixed effects. In the other cases, we anyway assume a linear approximation which proves to be effective in series of simulations while allowing us to maintain a simple approach.

Third, the logit formulation allows for a fixed-effects estimation approach based on sufficient statistics for the incidental parameters, thus avoiding parametric assumptions on the distribution of the individual unobserved heterogeneity. In particular, we propose estimating the model parameters by means of a Pseudo Conditional Maximum Likelihood (PCML) estimator recently put forward by Bartolucci and Nigro (2012), and here adapted to the proposed extended formulation. They approximate the dynamic logit with a Quadratic Exponential (QE) model (Cox, 1972; Bartolucci and Nigro, 2010), which admits a sufficient statistics for the incidental parameters and has the same interpretation as the dynamic logit model in terms of log-odds ratio between pairs of consecutive outcomes. In simpler contexts, this approach leads to a consistent estimator of the model parameters under the null hypothesis of absence of true state dependence, whereas has a reduced bias even with strong state dependence.

We study the finite sample properties of the PCML estimator for the proposed model through an extensive simulation study. The results show that the PCML estimator exhibits a negligible bias, for both the regression parameter associated with the predetermined covariate and the state dependence parameter, in the presence of substantial departures from noncausality. In addition, the estimation bias is almost negligible when the density of the predetermined covariate does not belong to the exponential family or its conditional mean depends on time-varying effects. It is also worth noting that the qualities of the proposed approach emerge for quite short T and a large number of cross-section units. Finally, the PCML is compared with the correlated random-effects ML estimator of Wooldridge (2005), adapted for the proposed formulation. This ML estimator is consistent for the parameters of interest in presence of feedbacks, although remarkably less efficient than the PCML in estimating the state dependence parameter, especially with short T. However, differently from our approach, consistency relies on the assumption of independence between the predetermined covariates and the individual unobserved effects, which is hardly tenable in practice.

The rest of the paper is organized as follows. Section 2 introduces the definitions of noncausality and strict exogeneity for nonlinear models. In Section 3 we illustrate the proposed model formulation. Section 4 describes the PCML estimation approach. Section

5 outlines the simulation study, and Section 6 provides main conclusions.

2 Definitions

Consider panel data for a sample of n units observed at T occasions according to a single explanatory variable x_{it} and binary response y_{it}, with $i = 1, \dots, n$ and $t = 1, \dots, T$, where the response variable is affected by a time-constant unobservable intercept c_i. Also let $\boldsymbol{x}_{i,t_1:t_2} = (x_{it_1}, \dots, x_{it_2})'$ and $\boldsymbol{y}_{i,t_1:t_2} = (y_{it_1}, \dots, y_{it_2})'$ denote the column vectors with elements referred to the period from the t_1-th to the t_2-th occasion, so that $\boldsymbol{x}_i = \boldsymbol{x}_{i,1:T}$ and $\boldsymbol{y}_i = \boldsymbol{y}_{i,1:T}$ are referred to the entire period of observation for the same sample unit i. Note that here we consider only one covariate to maintain the illustration simple, but all definitions and results below naturally extend to the case of more covariates per time occasion.

In this framework, and as illustrated in Chamberlain (1982), assuming that the economic life of any individual begins at time $t = 1$, the Granger's definition of noncausality is:

Definition. G - *The response (y) <u>does not cause</u> the covariate (x) conditional on the time-fixed effect (c) if $x_{i,t+1}$ is conditionally independent of $\boldsymbol{y}_{i,1:t}$, given c_i and $\boldsymbol{x}_{i,1:t}$, for all i and t, that is:*

$$p(x_{i,t+1}|c_i, \boldsymbol{x}_{i,1:t}, \boldsymbol{y}_{i,1:t}) = p(x_{i,t+1}|c_i, \boldsymbol{x}_{i,1:t}), \quad i = 1, \dots, n, \ t = 1, \dots, T-1. \tag{1}$$

Testing for G requires the knowledge and formulation of the model for each time-specific covariate given the the previous covariates and responses. However, following Chamberlain (1982), we introduce a condition that is the basis of the approach that we present in the next sections.

Definition. S' - *x is <u>strictly exogenous</u> with respect to y, given c and the past responses, if y_{it} is independent of $\boldsymbol{x}_{i,t+1:T}$ conditional on c_i, $\boldsymbol{x}_{i,1:t}$, and $\boldsymbol{y}_{i,1:t-1}$, for all i and t, that is*

$$p(y_{it}|c_i, \boldsymbol{x}_i, \boldsymbol{y}_{i,1:t-1}) = p(y_{it}|c_i, \boldsymbol{x}_{i,1:t}, \boldsymbol{y}_{i,1:t-1}), \quad i = 1, \dots, n, \ t = 1, \dots, T-1, \tag{2}$$

where $\boldsymbol{y}_{i,t-1}$ disappears from the conditioning argument for $t = 1$.

The following result holds, whose proof is related to that provided in Chamberlain (1982).

Theorem 1. G *and* S' *are equivalent conditions.*

Proof. G may be reformulated as

$$\frac{p(x_{i,t+1}, c_i, \boldsymbol{x}_{i,1:t}, \boldsymbol{y}_{i,1:t})}{p(c_i, \boldsymbol{x}_{i,1:t}, \boldsymbol{y}_{i,1:t})} = \frac{p(x_{i,t+1}, c_i, \boldsymbol{x}_{i,1:t})}{p(c_i, \boldsymbol{x}_{i,1:t})}, \quad t = 1, \dots, T-1,$$

for all i. Exchanging the denominator at lhs with the numerator at rhs, the previous equality becomes

$$p(\boldsymbol{y}_{i,1:t}|c_i, \boldsymbol{x}_{i,1:t+1}) = p(\boldsymbol{y}_{i,1:t}|c_i, \boldsymbol{x}_{i,1:t}), \quad t = 1, \dots, T-1,$$

which, by marginalization, implies that

$$p(\boldsymbol{y}_{i,1:s}|c_i, \boldsymbol{x}_{i,1:t+1}) = p(\boldsymbol{y}_{i,1:s}|c_i, \boldsymbol{x}_{i,1:t}), \quad t = 1, \dots, T-1, \ s = 1, \dots, t.$$

Therefore, we have

$$p(y_{is}|c_i, \boldsymbol{x}_{i,1:t+1}, \boldsymbol{y}_{i,1:s-1}) = p(y_{is}|c_i, \boldsymbol{x}_{i,1:t}, \boldsymbol{y}_{i,1:s-1}), \quad t = 1, \dots, T-1, \ s = 1, \dots, t.$$

Finally, by recursively using the previous expression for a fixed s and for t from $T-1$ to s we obtain condition S' as defined in (2). Similarly, S' implies that

$$p(\boldsymbol{x}_{i,t+1:T}|c_i, \boldsymbol{x}_{i,1:t}, \boldsymbol{y}_{i,1:t}) = p(\boldsymbol{x}_{i,t+1:T}|c_i, \boldsymbol{x}_{i,1:t}, \boldsymbol{y}_{i,1:t-1}), \quad t = 1, \dots, T-1,$$

for all i and implies

$$p(x_{i,s+1}|c_i, \boldsymbol{x}_{i,1:s}, \boldsymbol{y}_{i,1:t}) = p(x_{i,s+1}|c_i, \boldsymbol{x}_{i,1:s}, \boldsymbol{y}_{i,1:t-1}), \quad t = 1, \dots, T-1, \ s = 1, \dots, T-1,$$

which, in turn, leads to condition (1) and then G. □

It is worth noting that, apart from the case $T = 2$, definition S' is stronger than the definition of strict exogeneity of Sims (1972) adapted to the case of binary panel data, which we denote by S. Then, being equivalent to S', G implies S, but in general S does not imply G. In fact, S is expressed avoiding to condition on the previous responses:

Definition. S - *x is strictly exogenous with respect to y, given c, if y_{it} is independent of $\boldsymbol{x}_{i,t+1:T}$ conditional on c_i and $\boldsymbol{x}_{i,1:t}$, for all i and t, that is*

$$p(y_{it}|c_i, \boldsymbol{x}_i) = p(y_{it}|c_i, \boldsymbol{x}_{i,1:t}), \quad i = 1, \dots, n, \ t = 2, \dots, T. \tag{3}$$

Theorem 2. G *implies* S.

Proof. Proceeding as in the proof of Theorem 1, G implies that

$$p(y_{is}|c_i, \boldsymbol{x}_{i,1:t+1}) = p(y_{is}|c_i, \boldsymbol{x}_{i,1:t}), \quad t = 1, \dots, T-1, \ s = 1, \dots, t.$$

By recursively using the previous expression for a fixed s and for t from $T-1$ to s, we obtain condition (3). □

Although the focus here is on nonlinear binary panel data models, it is useful to accompany the discussion with the Granger's and the Sims' definitions in the simpler context of linear models, as laid out by Chamberlain (1984), where testable restrictions on the regression parameters can be derived directly. The starting point is a linear panel data model of the form

$$y_{it} = x_{it}\beta + c_i + \varepsilon_{it}, \quad i = 1, \ldots, n, \; t = 1, \ldots, T, \tag{4}$$

where now the dependent variables y_{it} are continuous and the error terms ε_{it} are iid. The usual exogeneity assumption is stated as

$$\mathrm{E}(\varepsilon_{it}|c_i, \boldsymbol{x}_i) = 0, \quad i = 1, \ldots, n, \; t = 1, \ldots, T, \tag{5}$$

which rules out the lagged response variables from the regression specification, as well as possible feedback effects from past values of y_{it} on to the present and future values of the covariate.

Now consider the minimum mean-square error linear predictor, denoted by $\mathrm{E}^*(\cdot)$, and consider the following definitions, which hold for all i:

$$\mathrm{E}^*(c_i|\boldsymbol{x}_i) = \eta + \boldsymbol{x}_i'\boldsymbol{\lambda}, \tag{6}$$

$$\mathrm{E}^*(y_{it}|\boldsymbol{x}_i) = \alpha_t + \boldsymbol{x}_i'\boldsymbol{\pi}_t, \quad t = 1, \ldots, T, \tag{7}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)'$ and $\boldsymbol{\pi}_t = (\pi_{t1}, \ldots, \pi_{tT})'$ are vectors of regression coefficients. Equation (7) may also be expressed as

$$\mathrm{E}^*(\boldsymbol{y}_i|\boldsymbol{x}_i) = \boldsymbol{\alpha} + \boldsymbol{\Pi}\boldsymbol{x}_i,$$

with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_T)'$ and $\boldsymbol{\Pi} = (\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_T)'$. It may be simply proved that assumptions (4), (5), together with definition (6), imply that

$$\boldsymbol{\Pi} = \beta\boldsymbol{I} + \boldsymbol{1}\boldsymbol{\lambda}',$$

where $\boldsymbol{I}$ is an identity matrix and $\boldsymbol{1}$ is a column vector of ones of suitable dimension; in the present case they are of dimension T. In Chamberlain (1984), the structure of $\boldsymbol{\Pi}$ is related to the definition of strict exogeneity in Sims (1972) for linear models (equivalent to condition S for binary models defined above) that, conditional on the permanent unobserved heterogeneity, is stated as

$$\mathrm{E}^*\left(y_{it}|c_i, \boldsymbol{x}_i\right) = \mathrm{E}^*\left(y_{it}|c_i, \boldsymbol{x}_{i,1:t}\right), \quad t = 1, \ldots, T. \tag{8}$$

Sims (1972) proved the equivalence of this condition with that of noncausality of Granger

(1969). In matrix notation, condition (8) can be written as

$$\mathrm{E}^*(\boldsymbol{y}_i|c_i, \boldsymbol{x}_i) = \boldsymbol{\varphi} + \boldsymbol{\Psi}\boldsymbol{x}_i + c_i\boldsymbol{\tau}, \tag{9}$$

where $\boldsymbol{\Psi}$ is a lower triangular matrix, $\boldsymbol{\tau} = (\tau_1, \dots, \tau_T)'$, and $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_T)'$. Assumptions (6) and (9) then imply the following structure for $\boldsymbol{\Pi}$:

$$\boldsymbol{\Pi} = \boldsymbol{B} + \boldsymbol{\delta}\boldsymbol{\lambda}',$$

where $\boldsymbol{B}$ is a lower triangular matrix and $\boldsymbol{\delta} = (\delta_1, \dots, \delta_T)'$.

It is straightforward to translate the restrictions in the structure of $\boldsymbol{\Pi}$ to the linear index function of a nonlinear model. In fact, Chamberlain (1984) and then Wooldridge (2010, Section 15.8.2) show that a simple test for strict exogeneity, S, in binary panel data models can be readily derived by adding $\boldsymbol{x}_{i,t+1}$ to the set of explanatory variables. In the next section we show not only that noncausality S' can be tested in a similar manner within a dynamic model formulation, but also that the linear index augmented with $\boldsymbol{x}_{i,t+1}$ represents, under rather general conditions, the exact log-odds ratio for the joint probability of y_{it} and $\boldsymbol{x}_{i,t+1}$ when S' is violated, thereby providing a model formulation that accounts for feedback effects and whose parameters may be consistently estimated.

3 Model formulation

Consider the general case in which, for $i = 1, \dots, n$ and $t = 1, \dots, T$, we observe a binary response variable y_{it} and a vector of k covariates denoted by $\boldsymbol{x}_{it}$. Then, we extend the previous notation by introducing $\boldsymbol{X}_{i,t_1:t_2} = (\boldsymbol{x}_{it_1}, \dots, \boldsymbol{x}_{it_2})$, with $\boldsymbol{X}_i = \boldsymbol{X}_{i,1:T}$ being the matrix of the covariates for all time occasions. In order to illustrate the proposed model, we first recall the main assumptions of the dynamic logit model.

3.1 Dynamic logit model

A standard formulation of a dynamic binary choice model assumes that, for $i = 1, \dots, n$ and $t = 1, \dots, T$, the binary response y_{it} has conditional distribution

$$p(y_{it}|c_i, \boldsymbol{X}_i, \boldsymbol{y}_{i,1:t-1}) = p(y_{it}|c_i, \boldsymbol{x}_{it}, y_{i,t-1}), \tag{10}$$

corresponding to a first-order Markov model for y_{it} with dependence only on the present values of the explanatory variables. The above conditioning set can be easily enlarged to include further lags of $\boldsymbol{x}_{it}$ and y_{it}.

Moreover, adopting a logit formulation for the conditional probability (see Hsiao, 2005, ch. 7, for a review), that is,

$$p(y_{it}|c_i, \boldsymbol{x}_{it}, y_{i,t-1}) = \frac{\exp\left[y_{it}\left(c_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + y_{i,t-1}\gamma\right)\right]}{1 + \exp\left(c_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + y_{i,t-1}\gamma\right)}, \quad t = 2, \ldots, T, \tag{11}$$

the conditional distribution of the overall vector of responses becomes:

$$p(\boldsymbol{y}_{i,2:T}|c_i, \boldsymbol{X}_i, y_{i1}) = \frac{\exp\left[y_{i+}c_i + \sum_{t=2}^{T} y_{it}\left(\boldsymbol{x}'_{it}\boldsymbol{\beta} + y_{it-1}\gamma\right)\right]}{\prod_{t=2}^{T}\left[1 + \exp\left(c_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + y_{i,t-1}\gamma\right)\right]}, \tag{12}$$

where $\boldsymbol{\beta}$ and γ are the parameters of interest for the covariates and the true state dependence (Heckman, 1981), respectively, $y_{i+} = \sum_{t=2}^{T} y_{it}$ is the *total score* and the individual-specific intercepts c_i are often considered as nuisance parameters; moreover, the initial observation y_{i1} is considered as given.

Expression (10) embeds assumption S' by excluding leads of $\boldsymbol{x}_{it}$ from the probability conditioning set. It therefore rules out feedbacks from the response variable to future covariates, that is, the Granger causality. Noncausality is often a hardly tenable assumption, as when the covariates of interest depend on individual choices. If covariates are predetermined, as opposed to strictly exogenous, estimation of the model parameters of interest can be severely biased, when estimation is based on eliminating or approximating c_i with quantities depending on the entire observed history of covariates (Mundlak, 1978; Chamberlain, 1984; Wooldridge, 2005).

3.2 Proposed model

As stated at the end of Section 2, dealing with violations of condition S', formulated as in (2), amounts to propose a generalization of the standard dynamic binary choice model based on assumption (10). In order to allow for such violations, we specify the probability of y_{it} conditional on individual intercept now denoted by d_i, $\boldsymbol{X}_i$, and $\boldsymbol{y}_{i,1:t-1}$ as

$$p(y_{it}|d_i, \boldsymbol{X}_i, \boldsymbol{y}_{i,1:t-1}) = p(y_{it}|d_i, \boldsymbol{X}_{i,t:t+1}, y_{i,t-1}), \tag{13}$$

retaining the assumption that previous covariates and responses before $y_{i,t-1}$ do not affect y_{it}. Note that, differently from (10), the conditioning set on the rhs includes the first-order leads of $\boldsymbol{x}_{it}$. Moreover, we use a different symbol for the unobserved individual intercept that, as will be clear in the following, is related to the individual parameter d_i. The formulation can easily be extended to include an arbitrary number of leads $\boldsymbol{X}_{i,t:t+H}$, with $H \leq T-3$, so that we retain at least two observations, which is necessary for inference (see Section 4). However, we do not explicitly consider this extension because, while being rather obvious, it strongly complicates the following exposition.[2] Following the discussion in Chamberlain (1984) and the suggestion in Wooldridge (2010, 15.8.2) on testing the strict exogeneity assumption, a test for noncausality can be derived by specifying the model as

$$p(y_{it}|d_i, \boldsymbol{X}_{i,t:t+1}, y_{i,t-1}) = g^{-1}(d_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{x}_{i,t+1}'\boldsymbol{\nu} + y_{i,t-1}\gamma), \quad t = 2, \ldots, T-1,$$

where $g^{-1}(\cdot)$ is an inverse link function. It is worth noting that the null hypothesis $H_0 : \boldsymbol{\nu} = 0$ corresponds to condition S', and then to Granger noncausality G. The identification of $\boldsymbol{\beta}$ and γ in presence of departures from noncausality requires further assumptions that lead to the formulation here proposed. In particular, we rely on the logit formulation

$$p(y_{it}|d_i, \boldsymbol{X}_{i,t:t+1}, y_{i,t-1}) = \frac{\exp\left[y_{it}\left(d_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{x}_{i,t+1}'\boldsymbol{\nu} + y_{it-1}\gamma\right)\right]}{1 + \exp\left(d_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{x}_{i,t+1}'\boldsymbol{\nu} + y_{i,t-1}\gamma\right)}. \tag{14}$$

Under a particular, very relevant, case this formulation is justified according to the following arguments.

First of all, denote the conditional density of the distribution of the covariate vector $\boldsymbol{x}_{i,t+1}$ as

$$f(\boldsymbol{x}_{i,t+1}|\boldsymbol{\xi}_i, \boldsymbol{X}_{i,1:t}, \boldsymbol{y}_{i,1:t}) = f(\boldsymbol{x}_{i,t+1}|\boldsymbol{\xi}_i, \boldsymbol{x}_{it}, y_{it}), \quad t = 1, \ldots, T-1, \tag{15}$$

where $\boldsymbol{\xi}_i$ is a column vector of time-fixed effects and the presence of y_{it} allows for feedback effects.[3] Then the logit for the distribution y_{it} conditional on c_i, $\boldsymbol{\xi}_i$, $\boldsymbol{X}_{i,t:t+1}$, and $y_{i,t-1}$ is

$$\log\frac{p(y_{it}=1|c_i, \boldsymbol{\xi}_i, \boldsymbol{X}_{i,t:t+1}, y_{i,t-1})}{p(y_{it}=0|c_i, \boldsymbol{\xi}_i, \boldsymbol{X}_{i,t:t+1}, y_{i,t-1})} = \log\frac{f(y_{it}=1, \boldsymbol{x}_{i,t+1}|c_i, \boldsymbol{\xi}_i, \boldsymbol{x}_{i,t}, y_{i,t-1})}{f(y_{it}=0, \boldsymbol{x}_{i,t+1}|c_i, \boldsymbol{\xi}_i, \boldsymbol{x}_{it}, y_{i,t-1})} =$$

$$\log\frac{p(y_{it}=1|c_i, \boldsymbol{x}_{it}, y_{i,t-1})f(\boldsymbol{x}_{i,t+1}|\boldsymbol{\xi}_i, \boldsymbol{x}_{it}, y_{it}=1)}{p(y_{it}=0|c_i, \boldsymbol{x}_{it}, y_{i,t-1})f(\boldsymbol{x}_{i,t+1}|\boldsymbol{\xi}_i, \boldsymbol{x}_{it}, y_{it}=0)}, \tag{16}$$

where the presence of time-fixed effects in the conditioning sets for y_{it} and $\boldsymbol{x}_{it}$ is determined by (13) and (15).[4] Furthermore, we assume that the probability of y_{it} conditional on c_i, $\boldsymbol{x}_{it}$, $y_{i,t-1}$ has the dynamic logit formulation expressed in (11) so that the above expression

[2]Chamberlain (1984) reports an empirical example where the linear index function of a logit model corresponds to the lhs of S in (3), where all the available lags and leads of $\boldsymbol{x}_{it}$ are used. However, this specification is valid only when $t = 1$ is the beginning of the subject's economic life. We do not make the same assumption here.

[3]In assumption (15) we maintain the same first-order dynamic as for (13). Nevertheless the assumptions on the conditioning set on the right-hand-side can be relaxed to include more lags of $\boldsymbol{x}_{it}$ and y_{it}.

[4]Notice that the extension of (13) to a number of leads $1 < H \leq T-3$ requires to rewrite the conditional density of covariates as $\prod_{h=1}^{H} r(\boldsymbol{x}_{i,t+h}|\boldsymbol{\xi}_i, \boldsymbol{x}_{i,t+h-1}, y_{it} = z)$, with $z = 0, 1$.

becomes

$$\log \frac{p(y_{it}=1|c_i,\boldsymbol{\xi}_i,\boldsymbol{X}_{i,t:t+1},y_{i,t-1})}{p(y_{it}=0|c_i,\boldsymbol{\xi}_i,\boldsymbol{X}_{i,t:t+1},y_{i,t-1})} = c_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + y_{i,t-1}\gamma + \log \frac{f(\boldsymbol{x}_{i,t+1}|\boldsymbol{\xi}_i,\boldsymbol{x}_{it},y_{it}=1)}{f(\boldsymbol{x}_{i,t+1}|\boldsymbol{\xi}_i,\boldsymbol{x}_{it},y_{it}=0)}.$$

The main point now is how to deal with the components involving the ratio between the conditional density of $\boldsymbol{x}_{i,t+1}$ for $y_{it}=0$ and $y_{it}=1$. Suppose that the conditional distribution of $\boldsymbol{x}_{i,t+1}$ belongs to the following exponential family:

$$f(\boldsymbol{x}_{i,t+1}|\boldsymbol{\xi}_i,\boldsymbol{x}_{it},y_{it}=z) = \frac{\exp[\boldsymbol{x}_{i,t+1}'(\boldsymbol{\xi}_i+\boldsymbol{\eta}_z)]h(\boldsymbol{x}_{i,t+1};\boldsymbol{\sigma})}{K(\boldsymbol{\xi}_i+\boldsymbol{\eta}_z;\boldsymbol{\sigma})}, \quad t=1,\dots,T-1,\ z=0,1, \tag{17}$$

where $h(\boldsymbol{x}_{i,t+1})$ is an arbitrary strictly positive function, possibly depending on suitable dispersion parameters $\boldsymbol{\sigma}$, and $K(\cdot)$ is the normalizing constant. Note that this structure also covers the case of $\boldsymbol{x}_{i,t+1}$ depending on time-fixed effects through $\boldsymbol{\xi}_i$. The following result holds, the proof of which is trivial.

Theorem 3. *Under assumptions* (11) *and* (17), *we have*

$$\log \frac{p(y_{it}=1|c_i,\boldsymbol{\xi}_i,\boldsymbol{X}_{i,t:t+1},y_{i,t-1})}{p(y_{it}=0|c_i,\boldsymbol{\xi}_i,\boldsymbol{X}_{i,t:t+1},y_{i,t-1})} = \log \frac{p(y_{it}=1|d_i,\boldsymbol{X}_{i,t:t+1},y_{i,t-1})}{p(y_{it}=0|d_i,\boldsymbol{X}_{i,t:t+1},y_{i,t-1})} =$$

$$d_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{x}_{i,t+1}'\boldsymbol{\nu} + y_{i,t-1}\gamma,$$

where $d_i = c_i + \log K(\boldsymbol{\xi}_i+\boldsymbol{\eta}_1;\boldsymbol{\sigma}) - \log K(\boldsymbol{\xi}_i+\boldsymbol{\eta}_0;\boldsymbol{\sigma})$ *and* $\boldsymbol{\nu} = \boldsymbol{\eta}_1 - \boldsymbol{\eta}_0$, *and then model* (14) *holds.*

Two cases satisfying (17) are for continuous covariates having multivariate normal distribution with common variance-covariance matrix and the case of binary covariates. More precisely, in the first case suppose that

$$\boldsymbol{x}_{i,t+1}|c_i,\boldsymbol{x}_{it},y_{it}=z \sim N(\boldsymbol{\zeta}_i+\boldsymbol{\mu}_z,\boldsymbol{\Sigma});$$

then (17) holds with $\boldsymbol{\xi}_i = \boldsymbol{\Sigma}^{-1}\boldsymbol{\zeta}_i$ and $\boldsymbol{\eta}_z = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_z$, $z=0,1$, where the upper (lower) triangular part of $\boldsymbol{\Sigma}$ go in $\boldsymbol{\psi}$. Regarding the second case, we suppose that given $\boldsymbol{\xi}_i$, $\boldsymbol{X}_{it}$, and $y_{it}=z$, the elements of $\boldsymbol{x}_{i,t+1}$ are conditionally independent, with the j-th element having Bernoulli distribution with success probability

$$\frac{\exp(\xi_{ij}+\eta_{zj})}{1+\exp(\xi_{ij}+\eta_{zj})}, \quad j=1,\dots,k,$$

where k is the number of covariates. In the other cases, when (17) does not hold, we anyway assume a linear approximation for the ratio between the conditional density of $\boldsymbol{x}_{i,t+1}$ for $y_{it}=0$ and $y_{it}=1$ in (16) which is the most natural solution to maintain an acceptable level of simplicity.

For the following developments, it is convenient to derive the conditional distribution of the entire vector of responses, which holds under the extended logit formulation (14) and that directly compares with (12). For all i, the distribution at issue is

$$p(\boldsymbol{y}_{i,2:T-1}|d_i, \boldsymbol{X}_i, y_{i1}, y_{iT}) = \tag{18}$$
$$\frac{\exp\left[y^*_{i+}d_i + \sum_{t=2}^{T-1} y_{it}\left(\boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu} + y_{it-1}\gamma\right)\right]}{\prod_{t=2}^{T-1}\left[1 + \exp\left(d_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu} + y_{i,t-1}\gamma\right)\right]}.$$

where $y^*_{i+} = \sum_{t=2}^{T-1} y_{it}$. In particular, model (18) reduces to the dynamic logit (12) under the null hypothesis of noncausality $H_0 : \boldsymbol{\nu} = \boldsymbol{0}$, if the probability in (12) is conditioned on y_{iT} and with different individual intercepts.

The parameters in (18) can be estimated by either a random- or fixed-effects approach, keeping in mind that a (correlated) random-effects strategy (Mundlak, 1978; Chamberlain, 1984) requires the predetermined covariates in $\boldsymbol{x}_{it}$ to be independent of d_i. As this assumption may often be hardly tenable, in the next section we discuss a fixed-effects estimation approach, first put forward by Bartolucci and Nigro (2012) and here adapted to the present case.

4 Fixed-effects estimation

With fixed-T panel data, a fixed-effects approach to the estimation of the parameters of the standard logit model is based on the maximization of the conditional likelihood given suitable sufficient statistics for the incidental parameters. The conditional estimator is common practice for static binary panel data models (Chamberlain, 1980), whereas, for the dynamic logit model, a sufficient statistic can only be derived in special cases: in absence of covariates with $T = 3$ (Chamberlain, 1985); with covariates on the basis of a weighted conditional log-likelihood, although the estimator is consistent only under certain conditions on the distribution of the covariates and the rate of convergence is slower than $\sqrt{n}$ (Honoré and Kyriazidou, 2000). These shortcomings have been overcome by Bartolucci and Nigro (2012), who approximate the dynamic logit with a QE model (Cox, 1972; Bartolucci and Nigro, 2010), which admits a sufficient statistic for the incidental parameters and has the same interpretation as the dynamic logit model in terms of log-odds ratio. Bartolucci and Nigro (2012) also propose to adopt a PCML estimator for the model parameters. In the following, we extend the approximating QE model to accommodate the parametrization of the proposed model formulation in (18).

4.1 Approximating model

The approximating model for (18) is derived by taking a linearization of the log-probability of the latter, similar to that used in Bartolucci and Nigro (2012), that is,

$$\log p(\boldsymbol{y}_{i,2:T-1}|d_i, \boldsymbol{X}_i, y_{i1}, y_{iT}) = y^*_{i+}d_i + \sum_{t=2}^{T-1} y_{it}\left(\boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu} + y_{i,t-1}\gamma\right) -$$

$$\sum_{t=2}^{T-1} \log\left[1 + \exp\left(d_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu} + y_{i,t-1}\gamma\right)\right]. \quad (19)$$

The term that is nonlinear in the parameters is approximated by a first-order Taylor series expansion around $d_i = \bar{d}_i$, $\boldsymbol{\beta} = \bar{\boldsymbol{\beta}}$, $\boldsymbol{\nu} = \bar{\boldsymbol{\nu}}$, and $\gamma = 0$, leading to

$$\sum_{t=2}^{T-1} \log\left[1 + \exp\left(d_i + \boldsymbol{x}'_{it}\bar{\boldsymbol{\beta}} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu} + y_{i,t-1}\gamma\right)\right] \approx$$

$$\sum_{t=2}^{T-1}\left[1 + \exp\left(\bar{d}_i + \boldsymbol{x}'_{it}\bar{\boldsymbol{\beta}} + \boldsymbol{x}'_{i,t+1}\bar{\boldsymbol{\nu}}\right)\right] +$$

$$\sum_{t=2}^{T-1} q_{it}\left[d_i - \bar{d}_i + \boldsymbol{x}'_{it}\left(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}\right) + \boldsymbol{x}'_{i,t+1}\left(\boldsymbol{\nu} - \bar{\boldsymbol{\nu}}\right)\right] + \sum_{t=2}^{T-1} q_{it}y_{i,t-1}\gamma, \quad (20)$$

where

$$q_{it} = \frac{\exp\left(\bar{d}_i + \boldsymbol{x}'_{it}\bar{\boldsymbol{\beta}} + \boldsymbol{x}'_{i,t+1}\bar{\boldsymbol{\nu}}\right)}{1 + \exp\left(\bar{d}_i + \boldsymbol{x}'_{it}\bar{\boldsymbol{\beta}} + \boldsymbol{x}'_{i,t+1}\bar{\boldsymbol{\nu}}\right)}.$$

Since only the last sum in (20) depends on $\boldsymbol{y}_{i,2:T-1}$, we can substitute (20) in (19) and obtain the approximation of the joint probability (18) that gives the following QE model

$$p^*(\boldsymbol{y}_{i,2:T-1}|d_i, \boldsymbol{X}_i, y_{i1}, y_{iT}) = \frac{\exp\left[y^*_{i+}d_i + \sum_{t=2}^{T-1} y_{it}\left(\boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu}\right) + \sum_t (y_{it} - q_{it})y_{i,t-1}\gamma\right]}{\sum_{\boldsymbol{z}_{2:T-1}} \exp\left[z^*_{+}d_i + \sum_{t=2}^{T-1} z_t\left(\boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu}\right) + \sum_t (z_t - q_{it})z_t\gamma\right]}, \quad (21)$$

where the sum at the denominator ranges over all the possible binary response vectors $\boldsymbol{z}_{2:T-1} = (z_2, \dots, z_{T-1})'$ and $z^*_+ = \sum_{t=2}^{T-1} z_t$, with $z_1 = y_{i1}$.

The joint probability in (21) is closely related to the probability of the response configuration $\boldsymbol{y}_{i,2:T-1}$ in the true model in (18). In particular, the approximating QE and the proposed true model share the properties summarized by the following theorem that can be proved along the lines of Bartolucci and Nigro (2010):[5]

[5] Results (*ii*) and (*iii*) can easily be derived by extending to the present case Theorem 1 in Bartolucci

Theorem 4. *For $i = 1, \dots, n$:*

(i) *In the case of $\gamma = 0$, the joint probability $p^*(\boldsymbol{y}_{i,2:T-1}|d_i, \boldsymbol{X}_i, y_{i1}, y_{iT})$ does not depend on $y_{i,t-1}$ or on q_{it}, and both the true (18) and approximating model (21), correspond to the following static logit model*

$$p^*(\boldsymbol{y}_{i,2:T-1}|d_i, \boldsymbol{X}_i, y_{i1}, y_{iT}) = \frac{\exp\left[y^*_{i+}d_i + \sum_{t=2}^{T-1} y_{it}\left(\boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu}\right)\right]}{\sum_{\boldsymbol{z}_{2:T-1}} \exp\left[z^*_+ d_i + \left(\boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu}\right)\right]} = \prod_{t=2}^{T-1} \frac{\exp\left[y_{it}\left(d_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu}\right)\right]}{1 + \exp\left(d_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu}\right)}.$$

(ii) *y_{it} is conditionally independent of $\boldsymbol{y}_{i,1:t-2}$ given d_i, $\boldsymbol{X}_i$, and $y_{i,t-1}$, for $t = 2, \dots, T$.*

(iii) *Under both models, the parameter γ has the same interpretation in terms of log-odds ratio between the responses y_{it} and $y_{i,t-1}$, for $t = 2, \dots, T-1$:*

$$\log \frac{p^*(y_{it} = 1|d_i, \boldsymbol{X}_i, y_{i,t-1} = 1)}{p^*(y_{it} = 0|d_i, \boldsymbol{X}_i, y_{i,t-1} = 1)} - \log \frac{p^*(y_{it} = 1|d_i, \boldsymbol{X}_i, y_{i,t-1} = 0)}{p^*(y_{it} = 0|d_i, \boldsymbol{X}_i, y_{i,t-1} = 0)} = \gamma.$$

The nice feature of the QE model in (21) is that it admits sufficient statistics for the incidental parameters d_i, which are the total scores y^*_{i+} for $i = 1, \dots, n$. The probability of $\boldsymbol{y}_{i,2:T-1}$, conditional on $\boldsymbol{X}_i$, y_{i1}, y_{iT}, and y^*_{i+}, for the approximating model is then

$$p^*\left(\boldsymbol{y}_{i,2:T-1}|\boldsymbol{X}_i, y_{i1}, y_{iT}, y^*_{i+}\right) = \frac{\exp\left[\sum_{t=2}^{T-1} y_{it}\left(\boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu}\right) + \sum_{t=2}^{T-1}(y_{it} - q_{it})y_{i,t-1}\gamma\right]}{\sum_{\substack{\boldsymbol{z}_{2:T-1}\\ z^*_+ = y^*_{i+}}} \exp\left[\sum_{t=2}^{T-1} z_t\left(\boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{x}'_{i,t+1}\boldsymbol{\nu}\right) + \sum_{t=2}^{T-1}(z_t - q_{it})z_{t-1}\gamma\right]}, \qquad (22)$$

which no longer depends on d_i and where the sum at the denominator is extended to all the possible response configurations $\boldsymbol{z}_{2:T-1}$ such that $z^*_+ = y^*_{i+}$, where $z^*_+ = \sum_{t=2}^{T-1}$.

4.2 Pseudo conditional maximum likelihood estimator

The formulation of the conditional log-likelihood for (22) relies on the fixed quantities q_{it}, that are based on a preliminary estimation of the parameters associated with the covariate and of the individual effects. Let $\boldsymbol{\phi} = (\boldsymbol{\beta}', \boldsymbol{\nu}')'$ be the vector collecting all the regression parameters and $\boldsymbol{\theta} = (\boldsymbol{\phi}', \boldsymbol{\gamma}')'$. The estimation approach is based on two-steps:

and Nigro (2012), that clarifies the connection between the QE and the dynamic logit model.

1. Preliminary estimates of the parameters needed to compute q_{it} are obtained by maximizing the following conditional log-likelihood

$$\begin{aligned}\ell(\bar{\boldsymbol{\phi}}) &= \sum_{i=1}^{n} 1\{0 < y_{it} < T-2\}\ell_i(\bar{\boldsymbol{\phi}}),\\ \ell_i(\bar{\boldsymbol{\phi}}) &= \log \frac{\exp\left[\sum_{t=2}^{T-1} y_{it}\left(\boldsymbol{x}_{it}'\bar{\boldsymbol{\beta}} + \boldsymbol{x}_{i,t+1}'\bar{\boldsymbol{\nu}}\right)\right]}{\sum_{\substack{\boldsymbol{z}_{2:T-1}\\ z_+^*=y_{i+}^*}} \exp\left[\sum_{t=2}^{T-1} z_t\left(\boldsymbol{x}_{it}'\bar{\boldsymbol{\beta}} + \boldsymbol{x}_{i,t+1}'\bar{\boldsymbol{\nu}}\right)\right]},\end{aligned}$$

   which can be maximized by a Newton-Raphson algorithm.

2. The parameter vector $\boldsymbol{\theta}$ is estimated by maximizing the conditional log-likelihood of (22), that can be written as

$$\begin{aligned}\ell^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}) &= \sum_i 1\{0 < y_{it} < (T-2)\}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}),\\ \ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}) &= \log p^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}(\boldsymbol{y}_{i,2:T-1}|\boldsymbol{X}_i, y_{i1}, y_{i1}, y_{i+}^*).\end{aligned} \tag{23}$$

   The resulting $\hat{\boldsymbol{\theta}}$ is the pseudo conditional maximum likelihood estimator.

Function $\ell^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$ may be maximized by Newton-Raphson using the score and observed information matrix reported below (Section 4.2.1). We also illustrate how to derive standard errors for the two-step estimator (Section 4.2.2). We leave out of the exposition the asymptotic properties of the PCML estimator, which can be derived along the same lines as in Bartolucci and Nigro (2012).

4.2.1 Score and information matrix

In order to write the score and information matrix for $\boldsymbol{\theta}$, it is convenient to rewrite $\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$ as

$$\begin{aligned}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}) = \; &\boldsymbol{u}^*(\boldsymbol{y}_{i,1:T-1})'\boldsymbol{A}^*(\boldsymbol{X}_i)'\boldsymbol{\theta} - \\ &\log \sum_{\substack{\boldsymbol{z}_{2:T-1}\\ z_+^*=y_{i+}^*}} \exp\left[\boldsymbol{u}^*(\boldsymbol{z}_{i,1:T-1})'\boldsymbol{A}^*(\boldsymbol{X}_i)'\boldsymbol{\theta}\right],\end{aligned} \tag{24}$$

where the notation $\boldsymbol{u}^*(\boldsymbol{y}_{i,1:T-1})$ is used to stress that $\boldsymbol{u}^*$ is a function of both the initial value y_{i1} and the response configuration $\boldsymbol{y}_{i,2:T-1}$; similarly $\boldsymbol{u}^*(\boldsymbol{z}_{i,1:T-1})$ is a function of y_{i1}

and $\boldsymbol{z}_{2:T-1}$, since $z_1 = y_{i1}$ as in (21). Moreover $\boldsymbol{u}^*(\boldsymbol{y}_{i,1:T-1})$ and $\boldsymbol{A}^*(\boldsymbol{X}_i)$ in (24) are

$$\begin{aligned}
\boldsymbol{u}^*(\boldsymbol{y}_{i,1:T-1}) &= \left(\boldsymbol{y}'_{i,2:T-1}, \sum_{t=2}^{T-1}(y_{it} - q_{it})y_{i,t-1}\right)' \\
\boldsymbol{A}^*(\boldsymbol{X}_i) &= \begin{pmatrix} \boldsymbol{X}_{i,2:T} & 0 \\ \boldsymbol{0}' & 1 \end{pmatrix},
\end{aligned} \tag{25}$$

where $\boldsymbol{X}_{i,2:T}$ is a matrix of $T-1$ rows and $2k$ columns, with k the number of covariates and typical row $\boldsymbol{x}'_{i,t:t+1}$, while $\boldsymbol{0}$ is column vector of zeros having a suitable dimension.[6] Using the above notation, the score $\boldsymbol{s}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}) = \nabla_{\boldsymbol{\theta}}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$ and the observed information matrix $\boldsymbol{J}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}) = -\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$ are

$$\begin{aligned}
\boldsymbol{s}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}) &= \sum_i 1\{0 < y_{i+}^* < T-2\}\boldsymbol{A}^*(\boldsymbol{X}_i)\{\boldsymbol{u}^*(\boldsymbol{y}_{i,2:T-1}) - \\
&\qquad \mathrm{E}^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}\left[\boldsymbol{u}^*(\boldsymbol{y}_{i,2:T-1})|\boldsymbol{X}_i, y_{i1}, , y_{iT}, y_{i+}^*\right]\},
\end{aligned} \tag{26}$$

and

$$\begin{aligned}
\boldsymbol{J}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}) &= \sum_i 1\{0 < y_{i+}^* < T-2\}\boldsymbol{A}^*(\boldsymbol{X}_i) \times \\
&\qquad \mathrm{V}^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}\left[\boldsymbol{u}^*(\boldsymbol{y}_{i,2:T-1})|\boldsymbol{X}_i, y_{i1}, y_{i+}^*\right]\boldsymbol{A}^*(\boldsymbol{X}_i)',
\end{aligned} \tag{27}$$

where the conditional expected value and variance are defined as

$$\begin{aligned}
&\mathrm{E}^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}\left[\boldsymbol{u}^*(\boldsymbol{y}_{i,2:T-1})|\boldsymbol{X}_i, y_{i1}, y_{i+}^*\right] = \\
&\qquad \sum_{\substack{\boldsymbol{z}_{H+1:T-H} \\ z_+^* = y_{i+}^*}} \boldsymbol{u}^*(\boldsymbol{z}_{i,2:T-2})p^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}\left(\boldsymbol{z}_{i,2:T-2}|\boldsymbol{X}_i, y_{i1}, y_{i+}^*\right),
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{V}^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}\left[\boldsymbol{u}^*(\boldsymbol{y}_{i,2:T-1})|\boldsymbol{X}_i, \boldsymbol{y}_{i,1:H}, y_{i+}^*\right] = \\
&\qquad \mathrm{E}^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}\left[\boldsymbol{u}^*(\boldsymbol{y}_{i,2:T-1})\boldsymbol{u}^*(\boldsymbol{y}_{i,2:T-1})'|\boldsymbol{X}_i, y_{i1}, y_{i+}^*\right] - \\
&\qquad \mathrm{E}^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}\left[\boldsymbol{u}^*(\boldsymbol{y}_{i,2:T-1})|\boldsymbol{X}_i, y_{i1}, y_{i+}^*\right]\mathrm{E}^*_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}\left[\boldsymbol{u}^*(\boldsymbol{y}_{i,2:T-1})|\boldsymbol{X}_i, y_{i1}, y_{i+}^*\right]'.
\end{aligned}$$

Following the results in Bartolucci and Nigro (2012), which can be applied directly to

[6] In order to clarify the structure of $\boldsymbol{A}^*(\boldsymbol{X}_i)$, consider the simple case of $T = 4$ time occasions and one covariate. Then

$$\boldsymbol{A}^*(\boldsymbol{X}_i) = \begin{pmatrix} x_{i2} & x_{i3} & 0 \\ x_{i3} & x_{i4} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

the present case, $\ell^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$ is always concave and $\boldsymbol{J}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$ is almost surely positive definite.[7] Then $\hat{\boldsymbol{\theta}}$ that maximizes $\ell^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$ is found at convergence of the standard Newton-Raphson algorithm.

4.2.2 Standard errors

The computation of standard errors must take into account the first step estimation of $\bar{\boldsymbol{\phi}}$. As Bartolucci and Nigro (2012) we also rely on the GMM approach (Hansen, 1982) and cast the estimating equations as

$$\boldsymbol{m}(\bar{\boldsymbol{\phi}},\boldsymbol{\theta}) = \sum_{i=1}^{n} 1\{0 < y_{i+}^* < T-2\}\boldsymbol{m}_i(\bar{\boldsymbol{\phi}},\boldsymbol{\theta}) = \boldsymbol{0},$$

where $\boldsymbol{m}_i(\bar{\boldsymbol{\phi}},\boldsymbol{\theta})$ contains the score vectors of the first step, $\nabla_{\bar{\boldsymbol{\phi}}}\ell_i(\bar{\boldsymbol{\phi}})$, and of the second step, $\nabla_{\boldsymbol{\theta}|\bar{\boldsymbol{\phi}}}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$. Then the GMM estimator is $(\tilde{\boldsymbol{\phi}}', \hat{\boldsymbol{\theta}}')'$ and its variance-covariance matrix can be estimated as

$$\boldsymbol{V}(\tilde{\boldsymbol{\phi}},\hat{\boldsymbol{\theta}}) = \boldsymbol{H}(\tilde{\boldsymbol{\phi}},\hat{\boldsymbol{\theta}})^{-1}\boldsymbol{S}(\tilde{\boldsymbol{\phi}},\hat{\boldsymbol{\theta}})\left[\boldsymbol{H}(\tilde{\boldsymbol{\phi}},\hat{\boldsymbol{\theta}})^{-1}\right]',$$

where

$$\begin{aligned}
\boldsymbol{S}(\bar{\boldsymbol{\phi}},\boldsymbol{\theta}) &= \sum_i 1\{0 < y_{i+}^* < T-2\}\boldsymbol{m}_i(\bar{\boldsymbol{\phi}},\boldsymbol{\theta})\boldsymbol{m}_i(\bar{\boldsymbol{\phi}},\boldsymbol{\theta})',\\
\boldsymbol{H}(\bar{\boldsymbol{\phi}},\boldsymbol{\theta}) &= \sum_i 1\{0 < y_{i+}^* < T-2\}\boldsymbol{H}_i(\bar{\boldsymbol{\phi}},\boldsymbol{\theta}).
\end{aligned}$$

Matrix $\boldsymbol{H}_i(\bar{\boldsymbol{\phi}},\boldsymbol{\theta})$ is composed of four blocks as follows:

$$\boldsymbol{H}_i(\bar{\boldsymbol{\phi}},\boldsymbol{\theta}) = \begin{pmatrix} \nabla_{\bar{\boldsymbol{\phi}}\bar{\boldsymbol{\phi}}}\ell_i(\bar{\boldsymbol{\phi}}) & \boldsymbol{0} \\ \nabla_{\boldsymbol{\theta}\bar{\boldsymbol{\phi}}}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}) & \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\beta}}) \end{pmatrix}.$$

The north-west block is expressed as

$$\nabla_{\bar{\boldsymbol{\phi}}\bar{\boldsymbol{\phi}}}\ell_i(\bar{\boldsymbol{\phi}}) = \boldsymbol{X}_{i,2:T}\mathrm{V}_{\bar{\boldsymbol{\phi}}}\left[\boldsymbol{u}(\boldsymbol{y}_{i,2:T-1})|\boldsymbol{X}_i, y_{i1}, y_{iT}, y_{i+}^*\right]\boldsymbol{X}_{i,2:T}',$$

where $\boldsymbol{X}_{i,2:T}$ is defined in (25) and $\mathrm{V}_{\bar{\boldsymbol{\phi}}}$ is the conditional variance in the static logit model. Moreover, $\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$ is equal to $-\boldsymbol{J}^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$; see definition (27). Finally, the derivation of $\nabla_{\boldsymbol{\theta}\bar{\boldsymbol{\phi}}}\ell_i^*(\boldsymbol{\theta}|\bar{\boldsymbol{\phi}})$ is not straightforward and we therefore rely on the numerical derivative of (26) with respect to $\bar{\boldsymbol{\phi}}$.

[7] See Bartolucci and Nigro (2012), Section 5, Theorem 2.

5 Simulation study

In this section we describe the design and illustrate the main results of the simulation study we used to investigate about the final sample properties of the PCML estimator for the parameters of the proposed model formulation. In the first part of the study, the main focus is on the performance under substantial departures from noncausality, which we obtain by a non-zero effect from the past values of the binary dependent variable on the present value of the covariate. In the second part, we compare the PCML estimator of (18) with an alternative ML random-effects estimator for the same model, based on the proposal by Wooldridge (2005) to account for the initial condition problem.

5.1 Simulation design

The simulation study is based on samples drawn from a dynamic logit model, where the linear index specification includes the lagged dependent variable, one explanatory variable x_{it} possibly predetermined, one strictly exogenous variable v_{it}, and individual unobserved heterogeneity. The model assumes that

$$y_{it} = 1\{c_i + \beta x_{it} - 0.5 v_{it} + \gamma y_{it-1} + \varepsilon_{it} \geq 0\}, \tag{28}$$

for $i = 1, \ldots, n$, $t = 2, \ldots, T$, with initial condition

$$y_{i1} = 1\{c_i + \beta x_{i1} - 0.5 v_{i1} + \varepsilon_{i1} \geq 0\}.$$

In the considered scenarios, the error terms ε_{it}, $t = 1, \ldots, T$, follow a logistic distribution with zero mean and variance equal to $\pi^2/3$ and the individual specific intercepts c_i are allowed to be correlated with x_{it} and v_{it}.

We consider a benchmark design and some extensions that are characterized by different choices for the distribution of the explanatory variable x_{it}. The general formulation is

$$\begin{aligned} x_{it} &= w(\xi_i + x^*_{it} + \psi v_{it} + \eta y_{it-1}), \\ x^*_{it} &\sim N(0, \pi^2/3), \end{aligned} \tag{29}$$

for $t = 2, \ldots, T$, the initial value is $x_{i1} = w(\xi_i + x^*_{i1} + \psi v_{i1})$ with x^*_{i1} being again a zero mean normal with variance $\pi^2/3$, and $v_{it} = \xi_i + v^*_{it}$, for $t = 1, \ldots, T$, where v^*_{it} is also $N(0, \pi^2/3)$. The parameter η governs the violation of S', stated in Section 2, and it takes value $\eta = 0$ under the assumption of noncausality, with $\eta \neq 0$ otherwise. In our benchmark design, we let $w(\cdot)$ be the identity function and $\psi = 0$, so that assumption (17) is satisfied and the model of Theorem 3 holds. We also consider two alternative designs where (17) does not hold and the model formulated in Theorem 3 is an approximation: first, we let $w(\cdot)$ be an indicator function so that x_{it} becomes a binary covariate with a normally distributed error term, with $p(x_{it}=1|\xi_i, v_{it}, y_{i,t-1}) = \Phi(\xi_i + x^*_{it} + \psi v_{it} + \eta y_{i,t-1})$, where $\Phi(\cdot)$ is the standard normal cdf and therefore does not belong to the exponential family in (17); secondly, we let the $w(\cdot)$ be the identity function and set $\psi = 0.5$ in order for x_{it} to depend on other time varying covariates.

Based on x^*_{it}, the individual intercepts c_i and ξ_i are derived as

$$\begin{aligned} c_i &= (1/T)\sum_{t=1}^{4} x^*_{it}, \\ \xi_i &= \varpi\, c_i + \sqrt{1-\varpi^2}u_{it}, \end{aligned} \tag{30}$$

with $\varpi = 0.5$, $u_{it} \sim N(0,1)$ and for $i = 1,\dots,n$. This way, the generating model admits a correlation between the covariates and the individual-specific intercepts and dependence between the unobserved heterogeneity in both processes for y and x.

In most economic applications, the parameters of interest are γ, measuring the state dependence, and the regression coefficient β. Based on the generating model (28), we ran experiments for scenarios with $\gamma = 0, 1$ and $\beta = 0, -1$. We examine violations of noncausality by setting $\eta = -1$, compared with the same scenarios with $\eta = 0$. The chosen values for β, γ, and η are consistent with likely situations in practice that relate, for instance, to the feedback effect of past employment on present child birth when analyzing female labor supply (see also Mosconi and Seri, 2006, for a related application). Notice that here we are implicitly assuming that the only source of contemporaneous endogeneity, namely the reverse causality between x_{it} and y_{it}, is completely captured by the correlation between the individual specific intercepts in the two processes. The sample sizes considered are $n = 500, 1000$ for $T = 4, 8$. The number of Monte Carlo replications is 1000.

5.2 Main results

Tables 1–6 report the main results of our simulation study. Tables 1–4 show the results for the benchmark design, under which the covariate x_{it}, generated as in (29), is normally distributed, with $w(\cdot)$ being the identity function, and $\psi = 0$, for all the combinations of the chosen values of β and γ. Tables 5 and 6 report the simulation results for the two extensions of our benchmark design, under which x_{it} is generated as a binary variable and with a dependence on the time varying covariate v_{it}, respectively, for $\beta = -1$, $\gamma = 1$, and $\eta = 0, -1$.

For each scenario, we investigate the finite sample performance of the PCML estimator in Section 4 for the proposed formulation (18) in two cases representing the null and alternative hypotheses of noncausality described by S' in Section 2: PCML$_1$ denotes the PCML estimator for the parameters in (18); PCML$_0$ denotes the estimator of (18) with the constraint $\nu = 0$. For each estimator, we report the mean bias, the median bias, the root-mean square error (RMSE), the median absolute error (MAE), as in Honoré and Kyriazidou (2000), and the t-tests at the 5% nominal size for $H_0 : \hat{\beta} = \beta$, and $H_0 : \hat{\gamma} = \gamma$. Finally we report the t-tests at the 5% nominal size for noncausality, $H_0 : \nu = 0$. We expect PCML$_0$ to yield biased estimators when $\eta \neq 0$ since, following S', the lead of x_{it} is omitted from the model specification. We limit the discussion to the estimation of β and γ, which are likely to be the parameters of main interest in applications. Results concerning the other model parameters are available upon request.

Table 1 summarizes the simulation results for our benchmark design and $\beta = \gamma = 0$. With $\eta = 0$, that is, in absence of feedback effects, the mean bias and median bias are always negligible, whereas the MAE and RMSE decrease with both n and T for the two models considered. While the same considerations hold for PCML$_1$ when $\eta = -1$, the PCML estimators of β provided by PCML$_0$ is severely biased and leads to misleading inference, although this pattern is alleviated for $T = 8$. The same patterns are shown in Table 2, where β is equal to -1. Moreover, the t-test for $H_0 : \nu = 0$ always attains its nominal size and exhibits strong empirical power in all the scenarios with $\eta = -1$

Tables 3 and 4 summarize simulation results for the same designs when $\gamma = 1$. They depict similar situations to those in Tables 1 and 2, with the exception of the bias of γ, that slightly increases. In fact, the performance of the PCML estimator may be especially sensitive to the degree of state dependence in the generated samples. A high value of γ leads to a reduction of the actual sample size via the indicator function in (23) and represents a large deviation from the approximating point by which (20) is derived. Nevertheless, Bartolucci and Nigro (2012) show that the bias and root-mean square error of PCML estimator of γ in the dynamic logit model decrease at a rate close to $\sqrt{n}$ and as T grows also for γ moving away from 0.

Tables 5 and 6 report the simulation results for two departures from the benchmark design: Table 5 refers to a binary covariate generated by a normal link function, while Table 6 refers to a normally distributed covariate depending on the time-varying covariate v_{it} (see Section 5.1 for details). These exercises are meant to investigate the properties of the PCML estimator when assumption (17) does not hold and the model formulated in Theorem 3 just embeds a linear approximation of (17). When the covariate is binary, the bias of the PCML$_1$ estimator of β and γ is always negligible. As for efficiency, the RMSE and MAE are slightly higher for β, although they decrease with both n and T (see Table 5). On the other hand, the results for $\psi = 0.5$ in Table 6 mirror closely those in Table 4, except for a larger bias with $T = 4$.

Table 1: Normally distributed covariate, $\beta = 0$, $\gamma = 0$, $\psi = 0$

	Estimation of β					Estimation of γ					$H_0 : \nu = 0$
	Mean bias	RMSE	Median bias	MAE	t-test	Mean bias	RMSE	Median bias	MAE	t-test	t-test
$\eta = 0$											
					$n = 500$, $T = 4$						
PCML$_1$	-0.003	0.072	-0.003	0.048	0.052	-0.026	0.305	-0.031	0.210	0.039	0.051
PCML$_0$	-0.001	0.060	0.001	0.039	0.045	-0.027	0.302	-0.025	0.209	0.036	
					$n = 500$, $T = 8$						
PCML$_1$	-0.000	0.027	0.000	0.018	0.066	0.003	0.106	0.002	0.073	0.055	0.037
PCML$_0$	-0.000	0.027	-0.000	0.018	0.062	0.003	0.106	0.002	0.073	0.056	
					$n = 1000$, $T = 4$						
PCML$_1$	0.000	0.051	-0.000	0.034	0.051	0.002	0.224	0.009	0.143	0.055	0.050
PCML$_0$	-0.000	0.043	-0.001	0.029	0.052	0.002	0.223	0.010	0.143	0.052	
					$n = 1000$, $T = 8$						
PCML$_1$	0.001	0.019	0.001	0.012	0.048	0.000	0.074	-0.002	0.048	0.053	0.055
PCML$_0$	0.001	0.018	0.001	0.012	0.053	0.000	0.074	-0.002	0.048	0.053	
$\eta = -1$											
					$n = 500$, $T = 4$						
PCML$_1$	0.002	0.078	-0.001	0.054	0.042	-0.013	0.338	-0.009	0.224	0.045	0.984
PCML$_0$	0.155	0.167	0.154	0.154	0.694	0.138	0.346	0.152	0.236	0.057	
					$n = 500$, $T = 8$						
PCML$_1$	-0.003	0.027	-0.002	0.018	0.047	-0.000	0.112	-0.000	0.076	0.044	1.000
PCML$_0$	0.048	0.054	0.048	0.048	0.498	0.053	0.115	0.049	0.078	0.078	
					$n = 1000$, $T = 4$						
PCML$_1$	-0.002	0.053	-0.002	0.037	0.051	-0.003	0.245	-0.002	0.166	0.055	1.000
PCML$_0$	0.149	0.155	0.149	0.149	0.935	0.149	0.275	0.153	0.195	0.089	
					$n = 1000$, $T = 8$						
PCML$_1$	-0.003	0.020	-0.004	0.014	0.071	0.004	0.080	0.003	0.055	0.046	1.000
PCML$_0$	0.048	0.052	0.048	0.048	0.795	0.057	0.092	0.056	0.063	0.129	

Table 2: Normally distributed covariate, $\beta = -1$, $\gamma = 0$, $\psi = 0$

	Estimation of β					Estimation of γ					$H_0 : \nu = 0$
	Mean bias	RMSE	Median bias	MAE	t-test	Mean bias	RMSE	Median bias	MAE	t-test	t-test
$\eta = 0$											
					$n = 500$,	$T = 4$					
PCML$_1$	-0.049	0.178	-0.027	0.106	0.039	0.037	0.482	0.028	0.325	0.056	0.055
PCML$_0$	-0.039	0.165	-0.020	0.102	0.048	0.033	0.473	0.018	0.318	0.056	
					$n = 500$,	$T = 8$					
PCML$_1$	-0.007	0.049	-0.005	0.034	0.057	-0.006	0.135	-0.000	0.095	0.049	0.045
PCML$_0$	-0.007	0.049	-0.004	0.033	0.056	-0.006	0.134	-0.001	0.094	0.053	
					$n = 1000$,	$T = 4$					
PCML$_1$	-0.019	0.117	-0.005	0.075	0.043	0.005	0.309	0.010	0.219	0.041	0.037
PCML$_0$	-0.015	0.111	-0.007	0.073	0.046	0.006	0.307	0.007	0.222	0.042	
					$n = 1000$,	$T = 8$					
PCML$_1$	-0.001	0.035	0.001	0.023	0.051	0.005	0.090	0.006	0.060	0.040	0.056
PCML$_0$	-0.001	0.035	0.001	0.022	0.055	0.005	0.090	0.007	0.060	0.041	
$\eta = -1$											
					$n = 500$,	$T = 4$					
PCML$_1$	-0.058	0.208	-0.037	0.122	0.058	-0.015	0.501	-0.020	0.333	0.051	0.808
PCML$_0$	0.122	0.199	0.138	0.158	0.222	0.045	0.474	0.058	0.317	0.050	
					$n = 500$,	$T = 8$					
PCML$_1$	-0.006	0.055	-0.005	0.035	0.049	0.002	0.148	0.002	0.101	0.058	1.000
PCML$_0$	0.047	0.069	0.048	0.052	0.194	-0.097	0.170	-0.095	0.122	0.112	
					$n = 1000$,	$T = 4$					
PCML$_1$	-0.027	0.134	-0.018	0.082	0.060	-0.003	0.340	-0.003	0.224	0.049	0.981
PCML$_0$	0.140	0.177	0.148	0.150	0.330	0.055	0.325	0.043	0.213	0.051	
					$n = 1000$,	$T = 8$					
PCML$_1$	-0.003	0.039	-0.003	0.027	0.056	0.007	0.101	0.007	0.069	0.055	1.000
PCML$_0$	0.050	0.061	0.051	0.051	0.311	-0.091	0.133	-0.091	0.096	0.172	

Table 3: Normally distributed covariate, $\beta = 0$, $\gamma = 1$, $\psi = 0$

	Estimation of β					Estimation of γ					$H_0 : \nu = 0$
	Mean bias	RMSE	Median bias	MAE	t-test	Mean bias	RMSE	Median bias	MAE	t-test	t-test
$\eta = 0$											
					$n = 500$, $T = 4$						
PCML$_1$	-0.002	0.079	0.001	0.051	0.033	-0.003	0.418	-0.000	0.289	0.063	0.040
PCML$_0$	-0.000	0.069	-0.003	0.046	0.025	-0.010	0.412	-0.013	0.288	0.060	
					$n = 500$, $T = 8$						
PCML$_1$	-0.002	0.027	-0.003	0.018	0.049	0.005	0.113	0.004	0.076	0.048	0.052
PCML$_0$	-0.002	0.027	-0.003	0.017	0.049	0.005	0.113	0.003	0.075	0.046	
					$n = 1000$, $T = 4$						
PCML$_1$	-0.003	0.054	-0.003	0.037	0.031	-0.025	0.279	-0.029	0.195	0.052	0.035
PCML$_0$	-0.002	0.048	-0.000	0.032	0.045	-0.029	0.277	-0.033	0.193	0.049	
					$n = 1000$, $T = 8$						
PCML$_1$	-0.001	0.020	-0.000	0.014	0.051	-0.001	0.080	-0.006	0.054	0.049	0.059
PCML$_0$	-0.001	0.020	-0.000	0.014	0.051	-0.002	0.080	-0.005	0.054	0.048	
$\eta = -1$											
					$n = 500$, $T = 4$						
PCML$_1$	0.006	0.085	0.008	0.056	0.037	-0.004	0.441	-0.016	0.297	0.050	0.894
PCML$_0$	0.143	0.157	0.143	0.143	0.520	0.147	0.442	0.140	0.281	0.055	
					$n = 500$, $T = 8$						
PCML$_1$	0.007	0.031	0.006	0.021	0.065	0.006	0.125	0.003	0.084	0.057	1.000
PCML$_0$	0.018	0.032	0.017	0.022	0.104	0.008	0.114	0.002	0.078	0.055	
					$n = 1000$, $T = 4$						
PCML$_1$	0.004	0.060	0.005	0.042	0.039	-0.001	0.301	-0.002	0.191	0.059	0.992
PCML$_0$	0.139	0.147	0.137	0.137	0.815	0.148	0.323	0.147	0.225	0.075	
					$n = 1000$, $T = 8$						
PCML$_1$	0.004	0.020	0.003	0.013	0.059	0.002	0.089	0.004	0.060	0.055	1.000
PCML$_0$	0.015	0.023	0.014	0.016	0.118	0.005	0.082	0.003	0.056	0.058	

Table 4: Normally distributed covariate, $\beta = -1$, $\gamma = 1$, $\psi = 0$

	Estimation of β					Estimation of γ					$H_0 : \nu = 0$
	Mean bias	RMSE	Median bias	MAE	t-test	Mean bias	RMSE	Median bias	MAE	t-test	t-test
$\eta = 0$											
					$n = 500$,	$T = 4$					
PCML$_1$	-0.030	0.207	0.003	0.120	0.056	0.056	0.571	0.032	0.365	0.038	0.056
PCML$_0$	-0.027	0.190	-0.001	0.106	0.052	0.045	0.560	0.035	0.360	0.038	
					$n = 500$,	$T = 8$					
PCML$_1$	-0.007	0.052	-0.005	0.036	0.048	0.005	0.143	0.006	0.092	0.059	0.056
PCML$_0$	-0.006	0.052	-0.003	0.036	0.048	0.005	0.142	0.004	0.093	0.055	
					$n = 1000$,	$T = 4$					
PCML$_1$	0.006	0.124	0.012	0.085	0.063	0.009	0.393	0.001	0.267	0.050	0.043
PCML$_0$	0.000	0.116	0.007	0.077	0.050	0.012	0.389	0.001	0.265	0.044	
					$n = 1000$,	$T = 8$					
PCML$_1$	-0.001	0.036	-0.001	0.024	0.047	0.009	0.100	0.011	0.064	0.057	0.058
PCML$_0$	-0.000	0.036	-0.000	0.024	0.047	0.009	0.099	0.009	0.065	0.056	
$\eta = -1$											
					$n = 500$,	$T = 4$					
PCML$_1$	-0.031	0.211	-0.002	0.133	0.045	0.035	0.632	0.032	0.392	0.041	0.509
PCML$_0$	0.123	0.219	0.148	0.175	0.185	0.053	0.590	0.055	0.386	0.045	
					$n = 500$,	$T = 8$					
PCML$_1$	-0.003	0.059	0.001	0.041	0.052	-0.020	0.158	-0.021	0.108	0.052	1.000
PCML$_0$	0.022	0.060	0.025	0.042	0.084	-0.150	0.211	-0.147	0.155	0.186	
					$n = 1000$,	$T = 4$					
PCML$_1$	0.012	0.139	0.025	0.095	0.057	0.018	0.405	0.012	0.269	0.035	0.809
PCML$_0$	0.151	0.193	0.165	0.168	0.334	0.045	0.391	0.042	0.261	0.037	
					$n = 1000$,	$T = 8$					
PCML$_1$	0.003	0.043	0.005	0.029	0.059	-0.016	0.113	-0.015	0.079	0.046	1.000
PCML$_0$	0.027	0.048	0.029	0.035	0.130	-0.145	0.180	-0.144	0.144	0.299	

Table 5: Binary covariate, $\beta = -1$, $\gamma = 1$, $\psi = 0$

	Estimation of β					Estimation of γ					$H_0 : \nu = 0$
	Mean bias	RMSE	Median bias	MAE	t-test	Mean bias	RMSE	Median bias	MAE	t-test	t-test
$\eta = 0$											
					$n = 500$, $T = 4$						
PCML$_1$	-0.007	0.352	-0.005	0.242	0.040	0.005	0.398	0.009	0.263	0.049	0.045
PCML$_0$	-0.011	0.309	0.001	0.210	0.038	-0.003	0.390	0.004	0.260	0.048	
					$n = 500$, $T = 8$						
PCML$_1$	-0.010	0.116	-0.010	0.078	0.050	0.000	0.113	-0.002	0.076	0.053	0.060
PCML$_0$	-0.008	0.115	-0.009	0.076	0.049	0.000	0.113	-0.001	0.076	0.051	
					$n = 1000$, $T = 4$						
PCML$_1$	0.019	0.238	0.023	0.160	0.042	-0.018	0.279	-0.029	0.187	0.060	0.045
PCML$_0$	-0.000	0.211	0.003	0.140	0.040	-0.019	0.277	-0.033	0.187	0.057	
					$n = 1000$, $T = 8$						
PCML$_1$	-0.009	0.080	-0.012	0.054	0.049	0.004	0.079	0.002	0.054	0.040	0.065
PCML$_0$	-0.008	0.079	-0.010	0.053	0.047	0.004	0.079	0.001	0.054	0.040	
$\eta = -1$											
					$n = 500$, $T = 4$						
PCML$_1$	0.022	0.364	0.038	0.236	0.044	0.001	0.409	-0.007	0.278	0.048	0.579
PCML$_0$	0.432	0.528	0.447	0.449	0.309	0.042	0.399	0.029	0.267	0.052	
					$n = 500$, $T = 8$						
PCML$_1$	0.008	0.121	0.005	0.083	0.047	-0.003	0.116	-0.009	0.080	0.048	1.000
PCML$_0$	0.049	0.124	0.046	0.080	0.074	-0.024	0.114	-0.027	0.083	0.049	
					$n = 1000$, $T = 4$						
PCML$_1$	0.044	0.265	0.063	0.185	0.048	-0.022	0.290	-0.032	0.193	0.052	0.883
PCML$_0$	0.447	0.494	0.450	0.451	0.553	0.029	0.283	0.018	0.189	0.055	
					$n = 1000$, $T = 8$						
PCML$_1$	0.013	0.088	0.014	0.057	0.063	-0.001	0.081	0.002	0.055	0.043	1.000
PCML$_0$	0.053	0.098	0.052	0.067	0.108	-0.022	0.081	-0.019	0.054	0.057	

Table 6: Normally distributed covariate, $\beta = -1$, $\gamma = 1$, $\psi = 0.5$

	Estimation of β					Estimation of γ					$H_0 : \nu = 0$
	Mean bias	RMSE	Median bias	MAE	t-test	Mean bias	RMSE	Median bias	MAE	t-test	t-test
$\eta = 0$											
					$n = 500$,	$T = 4$					
PCML$_1$	-0.075	0.278	-0.037	0.140	0.043	0.103	0.774	0.111	0.469	0.049	0.058
PCML$_0$	-0.044	0.222	-0.015	0.125	0.039	0.077	0.708	0.073	0.447	0.038	
					$n = 500$,	$T = 8$					
PCML$_1$	-0.006	0.058	-0.004	0.036	0.054	0.007	0.154	0.008	0.101	0.032	0.056
PCML$_0$	-0.004	0.057	-0.001	0.036	0.053	0.005	0.152	0.006	0.098	0.035	
					$n = 1000$,	$T = 4$					
PCML$_1$	-0.017	0.158	-0.009	0.099	0.064	0.013	0.491	-0.008	0.321	0.038	0.046
PCML$_0$	-0.008	0.144	0.004	0.091	0.063	0.009	0.475	-0.024	0.316	0.034	
					$n = 1000$,	$T = 8$					
PCML$_1$	-0.002	0.042	0.001	0.027	0.049	0.015	0.113	0.013	0.073	0.049	0.049
PCML$_0$	-0.001	0.041	0.001	0.027	0.047	0.015	0.112	0.013	0.074	0.051	
$\eta = -1$											
					$n = 500$,	$T = 4$					
PCML$_1$	-0.115	0.372	-0.045	0.170	0.062	0.087	0.970	0.022	0.527	0.071	0.408
PCML$_0$	0.092	0.257	0.132	0.184	0.164	0.059	0.810	0.008	0.475	0.065	
					$n = 500$,	$T = 8$					
PCML$_1$	-0.002	0.066	-0.001	0.044	0.057	-0.001	0.183	-0.000	0.119	0.061	1.000
PCML$_0$	0.027	0.067	0.028	0.048	0.092	-0.107	0.200	-0.101	0.133	0.115	
					$n = 1000$,	$T = 4$					
PCML$_1$	-0.027	0.191	-0.001	0.119	0.055	0.032	0.538	0.029	0.345	0.050	0.690
PCML$_0$	0.133	0.203	0.151	0.167	0.248	0.054	0.503	0.053	0.318	0.048	
					$n = 1000$,	$T = 8$					
PCML$_1$	0.001	0.046	0.002	0.032	0.060	-0.014	0.126	-0.014	0.084	0.055	1.000
PCML$_0$	0.029	0.053	0.030	0.037	0.121	-0.118	0.166	-0.119	0.124	0.173	

5.3 Comparison with alternative estimators

We compare the performance of the PCML estimator for model (18) with two alternative approaches. The first, denoted by W, is the correlated random-effects approach based on the proposal by Wooldridge (2005) for nonlinear dynamic panel data models, where the individual unobserved heterogeneity is assumed to be normally distributed and initial conditions are handled by specifying the distribution of c_i conditional on the initial value of $\boldsymbol{y}_i$. In Wooldridge (2005) a general formulation for this conditional distribution is proposed, where the individual random effects are allowed to depend on linear combinations of time-averages of strictly exogenous covariates (Mundlak, 1978). We specify the following conditional distribution of c_i

$$c_i|y_{i1} \sim y_{i1}\alpha + \bar{v}_i\pi + c_i^*, \qquad c_i^* \sim N(0, \sigma_c^2), \quad i = 1, \ldots, n.$$

where $\bar{v}_i = (1/T)\sum_{t=1}^{T} v_{it}$. It is worth noting that, in this case, the ML estimator of the model parameters is consistent if c_i^* is independent of the possibly predetermined covariate x_{it}. Therefore, we generate samples where c_i in (30) is distributed as a normal random variable with zero mean, unit variance, and $\varpi = 0$, in order to avoid the misspecification of the random effects. Nevertheless we also compare the ML and PCML estimator in the scenario where the individual intercepts are generated as in (30).

The second is the so-called infeasible logit estimator (Honoré and Kyriazidou, 2000) denoted by INF, where the generated individual intercepts are used as an additional covariate and the model parameters are then estimated by ML based on the pooled logit model formulation. The purpose is to compare the PCML estimator with a benchmark that is not sensitive to substantial deviations from the approximating model (20).

Tables 7 and 8 summarize the results of the simulation study, that we limit to the scenarios with $\beta = -1$, $\gamma = -1$, and $\eta = 0, -1$. Table 7 contains the results based on the samples generated with individual effects independent of the model covariate. The biases for β obtained by PCML and W are similar to those obtained by the infeasible logit, especially with $T = 8$, and the RMSE and MAE attain the same order of magnitude to those of INF with $n = 1000$ and $T = 8$. With $\eta = -1$, W shows a small bias for β and values of RMSE and MAE similar to the PCML estimator. As for γ, the bias of W increases with both values of η. This result is likely due to the fact that the actual number of time occasions exploited by the ML estimator is too small for W to deliver a negligible bias, for which at least 8 occasions are required (Akay, 2012). As expected, though, W exhibits rather large biases when the individual intercepts are generated as in (30) with $\varpi = 0.25$ (see Table 8), which makes the PCML a more attractive alternative since this is a scenario that is more likely to occur in practice.

Table 7: Normally distributed covariate, $\beta = -1$, $\gamma = 1$, $c_i \sim N(0, 1)$, $\varpi = 0$

	Estimation of β					Estimation of γ					$H_0 : \nu = 0$
	Mean bias	RMSE	Median bias	MAE	t-test	Mean bias	RMSE	Median bias	MAE	t-test	t-test
$\eta = 0$											
					$n = 500$, $T = 4$						
PCML	0.003	0.204	0.024	0.133	0.063	0.011	0.454	-0.006	0.317	0.051	0.059
W	-0.013	0.131	-0.003	0.090	0.054	0.030	0.279	0.031	0.199	0.045	0.046
INF	-0.012	0.094	-0.013	0.063	0.045	0.013	0.102	0.012	0.066	0.051	0.053
					$n = 500$, $T = 8$						
PCML	-0.003	0.064	0.002	0.045	0.049	0.005	0.121	0.004	0.080	0.047	0.046
W	-0.002	0.060	-0.000	0.041	0.045	0.005	0.112	0.005	0.076	0.045	0.040
INF	-0.002	0.055	-0.001	0.037	0.055	0.005	0.056	0.003	0.038	0.047	0.043
					$n = 1000$, $T = 4$						
PCML	0.022	0.138	0.032	0.092	0.052	-0.023	0.294	-0.031	0.197	0.045	0.060
W	-0.003	0.095	0.004	0.065	0.060	0.025	0.203	0.034	0.138	0.061	0.051
INF	-0.005	0.070	-0.003	0.047	0.064	0.007	0.069	0.007	0.047	0.042	0.037
					$n = 1000$, $T = 8$						
PCML	-0.000	0.046	-0.000	0.032	0.056	-0.003	0.083	-0.003	0.056	0.040	0.064
W	0.002	0.042	0.003	0.028	0.062	-0.003	0.079	-0.002	0.052	0.042	0.056
INF	0.002	0.038	0.001	0.024	0.057	0.002	0.040	0.002	0.028	0.041	0.044
$\eta = -1$											
					$n = 500$, $T = 4$						
PCML	-0.010	0.279	0.020	0.173	0.058	0.023	0.559	-0.004	0.343	0.049	0.993
W	-0.059	0.187	-0.035	0.118	0.044	0.061	0.370	0.072	0.254	0.059	1.000
INF	-0.014	0.113	-0.006	0.074	0.054	0.007	0.120	0.003	0.078	0.045	1.000
					$n = 500$, $T = 8$						
PCML	0.025	0.085	0.027	0.057	0.061	-0.030	0.162	-0.029	0.111	0.050	1.000
W	-0.051	0.090	-0.049	0.061	0.089	0.104	0.180	0.103	0.124	0.118	1.000
INF	-0.004	0.065	-0.001	0.043	0.047	0.002	0.070	-0.001	0.047	0.051	1.000
					$n = 1000$, $T = 4$						
PCML	0.016	0.182	0.028	0.128	0.048	-0.025	0.379	-0.028	0.242	0.044	1.000
W	-0.041	0.134	-0.036	0.087	0.059	0.063	0.268	0.062	0.186	0.060	1.000
INF	-0.010	0.081	-0.009	0.055	0.055	0.011	0.085	0.011	0.057	0.044	1.000
					$n = 1000$, $T = 8$						
PCML	0.025	0.064	0.026	0.044	0.077	-0.038	0.121	-0.038	0.080	0.071	1.000
W	-0.050	0.073	-0.050	0.053	0.148	0.096	0.143	0.098	0.105	0.165	1.000
INF	-0.000	0.046	0.000	0.030	0.052	0.003	0.050	0.001	0.033	0.061	1.000

Table 8: Normally distributed covariate, $\beta = -1$, $\gamma = 1$, $c_i = (1/T)\sum_{t=1}^{4} x_{it}^*$, $\varpi = 0.5$

	Estimation of β					Estimation of γ					$H_0 : \nu = 0$
	Mean bias	RMSE	Median bias	MAE	t-test	Mean bias	RMSE	Median bias	MAE	t-test	t-test
$\eta = 0$											
					$n = 500$, $T = 4$						
PCML	-0.002	0.194	0.018	0.123	0.053	-0.011	0.432	-0.020	0.290	0.055	0.057
W	0.162	0.205	0.167	0.169	0.311	-0.205	0.362	-0.197	0.245	0.085	0.663
INF	-0.011	0.098	-0.008	0.066	0.058	0.016	0.094	0.015	0.064	0.049	0.065
					$n = 500$, $T = 8$						
PCML	-0.011	0.065	-0.009	0.045	0.052	0.005	0.118	0.003	0.078	0.041	0.054
W	0.056	0.082	0.058	0.061	0.183	-0.061	0.125	-0.064	0.089	0.067	0.277
INF	-0.005	0.054	-0.006	0.036	0.051	0.005	0.056	0.005	0.038	0.055	0.050
					$n = 1000$, $T = 4$						
PCML	0.028	0.132	0.038	0.094	0.058	-0.010	0.305	-0.009	0.206	0.056	0.051
W	0.173	0.194	0.176	0.176	0.547	-0.196	0.289	-0.193	0.212	0.148	0.915
INF	-0.005	0.068	-0.006	0.045	0.059	0.006	0.067	0.009	0.045	0.058	0.053
					$n = 1000$, $T = 8$						
PCML	-0.003	0.043	-0.004	0.028	0.047	0.006	0.088	0.004	0.057	0.058	0.045
W	0.063	0.074	0.062	0.062	0.325	-0.060	0.100	-0.060	0.070	0.124	0.488
INF	-0.000	0.037	-0.000	0.025	0.037	0.000	0.039	-0.000	0.026	0.051	0.037
$\eta = -1$											
					$n = 500$, $T = 4$						
PCML	0.002	0.276	0.021	0.177	0.058	0.003	0.534	-0.023	0.356	0.039	0.996
W	0.057	0.200	0.072	0.143	0.101	0.007	0.416	0.037	0.293	0.060	1.000
INF	-0.068	0.130	-0.068	0.083	0.074	0.229	0.255	0.226	0.226	0.517	1.000
					$n = 500$, $T = 8$						
PCML	0.023	0.086	0.021	0.060	0.059	-0.030	0.163	-0.033	0.114	0.055	1.000
W	-0.020	0.079	-0.016	0.055	0.062	0.060	0.158	0.058	0.107	0.073	1.000
INF	-0.016	0.069	-0.016	0.046	0.056	0.117	0.135	0.116	0.116	0.399	1.000
					$n = 1000$, $T = 4$						
PCML	0.023	0.183	0.040	0.122	0.050	-0.022	0.370	-0.028	0.245	0.045	1.000
W	0.072	0.152	0.075	0.107	0.129	0.007	0.298	0.011	0.203	0.066	1.000
INF	-0.063	0.102	-0.060	0.070	0.114	0.222	0.236	0.220	0.220	0.814	1.000
					$n = 1000$, $T = 8$						
PCML	0.024	0.062	0.025	0.042	0.075	-0.029	0.115	-0.032	0.084	0.059	1.000
W	-0.021	0.057	-0.022	0.039	0.056	0.057	0.116	0.058	0.081	0.076	1.000
INF	-0.018	0.050	-0.017	0.033	0.061	0.113	0.123	0.113	0.113	0.669	1.000

6 Conclusions

In this paper, we propose a novel model formulation for dynamic binary panel data models that accounts for feedback effects from the past of the outcome variable on the present value of covariates. Our proposal is particularly well suited for short panels with a large number of cross-section units, typically provided by rotated or strongly unbalanced continuous surveys, often employed for microeconomic applications. Our formulation is based on the equivalence between Granger's definition of noncausality and a modification of the Sims' strict exogeneity assumption for nonlinear panel data models, introduced by Chamberlain (1982) and for which we provide a more general theorem.

Under the logit model, the proposed model formulation yields three main advantages compared to the few available alternatives: (*i*) it does not require the specification of a parametric model for the predetermined explanatory variables; (*ii*) it has a simple formulation and allows, in practice, for the inclusion of a large number of predetermined covariates, discrete or continuous; (*iii*) its parameters can be estimated within a fixed-effects approach by a PCML, thereby allowing for an arbitrary dependence structure between the model covariates and the individual permanent unobserved heterogeneity.

From our simulation results, it emerges that PCML provides consistent estimation of the regression and state dependence parameters in presence of substantial departures from noncausality and that the bias is negligible even when the conditions for the exact logit model formulation are violated. Furthermore, we show that the alternative random-effects ML estimator based on Wooldridge (2005) for the model here proposed exhibits comparable finite-sample properties, provided the dependence between the predetermined covariate and the unobserved heterogeneity is reliably accounted for.

Finally, the logit model here proposed is fairly easy to estimate using available software. The PCML estimator of the proposed model can be implemented using the package `cquad` (Bartolucci and Pigini, 2016), whereas any routine for the random-effects logit model can be used for correlated-random effects ML estimator.

References

Akay, A. (2012). Finite-sample comparison of alternative methods for estimating dynamic panel data models. *Journal of Applied Econometrics*, 27(7):1189–1204.

Alessie, R., Hochguertel, S., and van Soest, A. (2004). Ownership of Stocks and Mutual Funds: A Panel Data Analysis. *The Review of Economics and Statistics*, 86(3):783–796.

Anderson, T. W. and Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American statistical Association*, 76(375):598–606.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies*, 58(2):277–297.

Arellano, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of econometrics*, 68(1):29–51.

Arellano, M. and Carrasco, R. (2003). Binary choice panel data models with predetermined variables. *Journal of Econometrics*, 115(1):125–157.

Arulampalam, W. (2002). State dependence in unemployment incidence: evidence for british men revisited. Technical report, IZA Discussion paper series.

Barndorff-Nielsen, O. (1978). *Information and exponential families in statistical theory*. John Wiley & Sons.

Bartolucci, F. and Nigro, V. (2010). A dynamic model for binary panel data with unobserved heterogeneity admitting a $\sqrt{n}$-consistent conditional estimator. *Econometrica*, 78:719–733.

Bartolucci, F. and Nigro, V. (2012). Pseudo conditional maximum likelihood estimation of the dynamic logit model for binary panel data. *Journal of Econometrics*, 170:102–116.

Bartolucci, F. and Pigini, C. (2016). `cquad`: An R and Stata package for conditional maximum likelihood estimation of dynamic binary panel data models. *Journal of Statistical Software*, In press.

Bettin, G. and Lucchetti, R. (2016). Steady streams and sudden bursts: persistence patterns in remittance decisions. *Journal of Population Economics*, 29(1):263–292.

Biewen, M. (2009). Measuring state dependence in individual poverty histories when there is feedback to employment status and household composition. *Journal of Applied Econometrics*, 24(7):1095–1116.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*, 87(1):115–143.

Brown, S., Ghosh, P., and Taylor, K. (2014). The existence and persistence of household financial hardship: A bayesian multivariate dynamic logit framework. *Journal of Banking & Finance*, 46:285–298.

Cappellari, L. and Jenkins, S. P. (2004). Modelling low income transitions. *Journal of Applied Econometrics*, 19(5):593–610.

Carrasco, R. (2001). Binary choice with binary endogenous regressors in panel data. *Journal of Business & Economic Statistics*, 19(4):385–394.

Carro, J. M. and Traferri, A. (2012). State dependence and heterogeneity in health using a bias-corrected fixed-effects estimator. *Journal of Applied Econometrics*, 29(2):181–207.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The Review of Economic Studies*, 47(1):225–238.

Chamberlain, G. (1982). The general equivalence of granger and sims causality. *Econometrica: Journal of the Econometric Society*, 50(3):569–581.

Chamberlain, G. (1984). Panel data. *Handbook of Econometrics*, 2:1247–1318.

Chamberlain, G. (1985). Heterogeneity, omitted variable bias, and duration dependence. In Heckman, J. J. and Singer, B., editors, *Longitudinal Analysis of Labor Market Data*. Cambridge University Press: Cambridge.

Contoyannis, P., Jones, A. M., and Rice, N. (2004). Simulation-based inference in dynamic panel probit models: an application to health. *Empirical Economics*, 29(1):49–77.

Cox, D. (1972). The analysis of multivariate binary data. *Applied statistics*, 21(2):113–120.

Florens, J.-P. and Mouchart, M. (1982). A note on noncausality. *Econometrica: Journal of the Econometric Society*, 50(3):583–591.

Giarda, E. (2013). Persistency of financial distress amongst italian households: Evidence from dynamic models for binary panel data. *Journal of Banking & Finance*, 37(9):3425 – 3434.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 37(3):424–438.

Halliday, T. J. (2008). Heterogeneity, state dependence and health. *The Econometrics Journal*, 11(3):499–516.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, 50(4):1029–1054.

Heckman, J. J. (1981). Heterogeneity and state dependence. *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press: Cambridge MA. Manski CF, McFadden (eds).

Heckman, J. J. and Borjas, G. J. (1980). Does unemployment cause future unemployment? definitions, questions and answers from a continuous time model of heterogeneity and state dependence. *Economica*, 47(187):pp. 247–283.

Heiss, F. (2011). Dynamics of self-rated health and selective mortality. *Empirical economics*, 40(1):119–140.

Honoré, B. E. and Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68(4):839–874.

Honoré, B. E. and Lewbel, A. (2002). Semiparametric binary choice panel data models without strictly exogeneous regressors. *Econometrica*, 70(5):2053–2063.

Hsiao, C. (2005). *Analysis of Panel Data*. Cambridge University Press, New York, 2nd edition.

Hyslop, D. R. (1999). State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women. *Econometrica*, 67(6):1255–1294.

Keane, M. P. and Sauer, R. M. (2009). Classification error in dynamic discrete choice models: Implications for female labor supply behavior. *Econometrica*, 77(3):975–991.

Michaud, P.-C. and Tatsiramos, K. (2011). Fertility and female employment dynamics in europe: the effect of using alternative econometric modeling assumptions. *Journal of Applied Econometrics*, 26(4):641–668.

Mosconi, R. and Seri, R. (2006). Non-causality in bivariate binary time series. *Journal of Econometrics*, 132(2):379–407.

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*, 46(1):69–85.

Pigini, C., Presbitero, A. F., and Zazzaro, A. (2016). State dependence in access to credit. *Journal of Financial Stability*, pages –. forthcoming.

Sims, C. A. (1972). Money, income, and causality. *The American Economic Review*, 62(4):540–552.

Stewart, M. B. (2007). The interrelated dynamics of unemployment and low-wage employment. *Journal of Applied Econometrics*, 22(3):511–531.

Wooldridge, J. M. (2000). A framework for estimating dynamic, unobserved effects panel data models with possible feedback to future explanatory variables. *Economics Letters*, 68(3):245–250.

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, 20(1):39–54.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. The MIT press.

Wunder, C. and Riphahn, R. T. (2014). The dynamics of welfare entry and exit amongst natives and immigrants. *Oxford Economic Papers*, 66(2):580–604.